

# **Grease-II Workshop**

## **The GREASE Project**

Mário J. Gaspar da Silva

Universidade de Lisboa, Faculdade de Ciências,  
Departamento de Informática

[mjs@di.fc.ul.pt](mailto:mjs@di.fc.ul.pt)

# **XLDB Group @ LASIGE:**

## **Bird's eye view**

- **LASIGE**: large-scale information systems lab
- **Research Lines**
  - Biomedical Informatics
  - Information Management(overlaps:  
epidemiology,  
health search,  
text mining, ...)



# Information Management Research Line @XLDB

- Systems for
  - data analysis
  - information integration
  - user access to large/  
complex datafrom heterogeneous  
platforms.
- Current research topics
  - Search: geographic  
information retrieval
  - Text mining and NLP
  - Web archiving
  - Information visualization

# GREASE (until 2006)

- ☐ Computational linguistics-based methods for reasoning about the geographic scope of web pages, using the Portuguese language Web as a test-bed;
- ☒ Web linkage analysis for assigning geographic scopes to Web pages;
- ☒ Information retrieval algorithms for matching users location with geographic scopes of Web pages;
- ☐ The design of a Semantic Web search appliance to manage users preferences data for query input and presentation of search results grouped by geographic proximity.

# **GREASE Contribs**

- Web mining sw for geographic names and geographic relationships
- GKB
  - Geo-Ontology creation + Ontology enrichment
- Geo similarity metrics
- Visualisation: mobile search results

# GeoCLEF Participations

- 2006 – lousy
- 2007 – top
  - BM25 tuned + manually tuned mixture of similarities
    - $\text{Final\_sim}(q,d) = b * \text{Geo\_sim}(q,d) + (1-b) \text{BM25sim}(q,d)$

# GREASE-II

- Methods for automatic generation of **geographic summaries** for documents
- **Query expansion** and automatic relevance feedback methods for geographic queries.
- **Fusion** of geographic and textual **rankings**.
- Support for **multi-lingual documents** and a **model of the physical geography** of the world in the knowledge-base.
- Information selection from large data sets using geographic criteria through **faceted interfaces**.



# Grease-II WP

1. Project Management
2. Multi-lingual geographic terminologies and Ontologies
3. **Geographic Information Extraction and Knowledge Integration**
4. **Geographic Information Retrieval: geographicity, geographic pseudo-relevance feedback, geo-similarity.**
5. **Faceted interfaces for GIR**
6. Prototyping
7. Evaluation and Dissemination

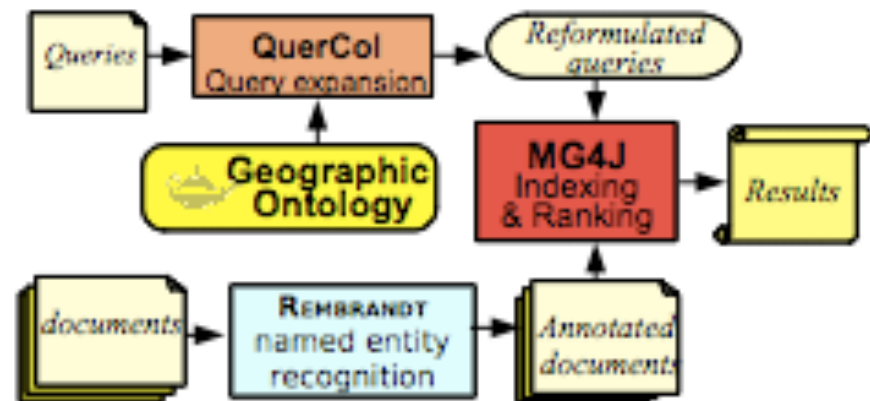
# Geo-Ontology

- GKB model now extended with physical domain  
(rivers, mountains, landmarks, ...)
- Language support
  - Multi-lingual
  - Demonyms (Welsh, Finn)
  - Connecting words (*do* Porto)
- ***GEO-NET-PT-02 “about”*** to be released
  - [http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT\\_02](http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02)



# GeoClef Participation

- 2008 – now second to Saarbrücken
  - QR + **geo terms reweighted in BM25 formula**
  - Introduced **implicit geo-names**
- Grease prototype now Buggy and clunky



# GREASE-II Calendar

- Submitted 8/2006
- Approved 8/2007
- “Started” 1/2008
  - New students 09/2008
- “Ends” 12/2009
  - Will likely request extension until 09/2010

# Outline

- 14:30 Introduction
  - 15:00 Geo Similarity
  - 15:15 Geographic Summaries
  - 15:30 Geo-aware advertisement selection
  - 15:45 Faceted interfaces for geographic data
  - 16:00 demo
- <break>
- 16:30 -17:00 Grease team discussion